Modeling and Simulation 建模与仿真, 2026, 15(4), 181-192
Published Online April 2026 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2026.154063

Hans汉斯

SegNext框架下融合边缘特征的息肉分割方法

刘一帆，魏　赟*

上海理工大学光电信息与计算机工程学院，上海

收稿日期：2026年3月27日；录用日期：2026年4月22日；发布日期：2026年4月27日

摘　要

结直肠息肉的精准分割是结直肠癌早期诊断与治疗的关键技术环节，现有融合卷积神经网络(Convolutional Neural Network, CNN)与视觉Transformer (Vision Transformer, ViT)的双分支架构，在息肉医学图像分割中仍面临特征融合不充分、边缘细节捕捉不足及语义表征鲁棒性欠佳等问题。针对上述问题，本文提出一种基于SegNext框架的多模态边缘感知提示息肉分割方法。该方法以多模态边缘感知适配器(Multi-modal Edge-Aware Adapter, MEAA)为核心，协同提取CNN局部纹理特征、小波高频分量(Wavelet High-Frequency Component, WHFC)边缘特征及ViT全局语义特征，通过通道－空间双注意力机制实现多模态特征的自适应融合与增强，并引入密集图视觉提示策略以弥补细节丢失。在Kvasir-SEG、CVC-ClinicDB等5个公开息肉分割数据集上的验证实验表明，所提框架在0次、1次、2次交互提示设置下，平均Dice系数分别达到0.854、0.910、0.935，平均交并比(Intersection over Union, IoU)分别达到0.781、0.858、0.895，分割性能显著优于现有主流方法，具备临床实时应用潜力，为结直肠息肉计算机辅助诊断提供了高效精准的技术方案。

关键词

息肉分割，预训练模型，微调，交互提示，多模态特征融合

A Polyp Segmentation Method Fusing Edge Features under the SegNext Framework

Yifan Liu, Yun Wei*

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: March 27, 2026; accepted: April 22, 2026; published: April 27, 2026

*通讯作者。

文章引用: 刘一帆, 魏赟. SegNext 框架下融合边缘特征的息肉分割方法[J]. 建模与仿真, 2026, 15(4): 181-192.
DOI: 10.12677/mos.2026.154063

Abstract

Accurate segmentation of colorectal polyps is a key technical step in the early diagnosis and treatment of colorectal cancer. Existing dual-branch architectures that combine Convolutional Neural Networks (CNN) and Vision Transformers (ViT) still suffer from insufficient feature fusion, inadequate capture of edge details, and poor robustness of semantic representation in polyp medical image segmentation. To address these issues, this paper proposes a multi-modal edge-aware prompt polyp segmentation method based on the SegNext framework. The method is centered on a Multi-modal Edge-Aware Adapter (MEAA), which jointly extracts local texture features from CNN, edge features from Wavelet High-Frequency Components (WHFC), and global semantic features from ViT. A channel-spatial dual attention mechanism is adopted to achieve adaptive fusion and enhancement of multi-modal features, and a dense map visual prompt strategy is introduced to compensate for detail loss. Experimental results on five public polyp segmentation datasets, including Kvasir-SEG and CVC-ClinicDB, show that the proposed framework achieves mean Dice coefficients of 0.854, 0.910, and 0.935, and mean Intersection over Union (IoU) scores of 0.781, 0.858, and 0.895 under 0, 1, and 2 interactive prompt settings, respectively. The segmentation performance is significantly superior to existing mainstream methods, demonstrating potential for real-time clinical application and providing an efficient and accurate technical solution for computer-aided diagnosis of colorectal polyps.

Keywords

Polyp Segmentation, Pre-Trained Model, Fine-Tuning, Interactive Prompts, Multi-Modal Feature Fusion

1. 引言

结直肠息肉的精准分割是结直肠癌早期诊断与治疗的关键，对提升筛查效率、改善诊疗预后意义重大。近年来，深度学习推动医学图像分割革新，UNet 及其衍生变体[1]-[5]凭借高效局部特征提取成为主流，但受 CNN 感受野局限，应对形态复杂、边界模糊的息肉时性能仍有提升空间[6]-[9]。融合 CNN 与 Transformer 的双分支架构[10]-[13]虽结合局部建模与全局表征优势，却因特征融合策略简单，未能充分挖掘异构特征互补价值。

Segment Anything Model (SAM) [14]-[17]通过交互式提示实现灵活分割，但稀疏向量表示易丢失息肉边缘细节；SegNext [18]引入密集图保留空间特征，缓解了细节丢失问题，却未针对息肉形态特点优化，难以精准提取局部纹理与边缘特征，无法满足临床高精度需求。

针对当前方法存在的全局与局部特征融合不充分、边缘细节捕捉薄弱等问题，同时弥补 SegNext 针对性缺陷，本文提出基于 SegNext 框架的边缘感知提示息肉分割方法。该方法整合三类异构特征流：CNN 中间层特征(捕捉局部细节与语义)、小波高频分量(WHFC，编码边缘先验)、ViT 补丁嵌入特征(刻画全局空间关系)；设计多模态边缘感知适配器(MEAA)，通过通道-空间双注意力机制实现特征自适应融合，生成高质量提示特征并嵌入 ViT 分支；结合 SegNext 密集图提示策略，强化空间细节与高层语义，全方位挖掘息肉特征，提升 SegNext 领域适应性，实现结直肠息肉鲁棒精准分割。

2. 模型结构设计

如图 1 所示，给定输入息肉图像 I，CNN 分支通过多阶段编码器逐步提取多级特征 $F_{cnn}^{i}\left(i=1,2,3,4\right)$。其中浅层特征 F_{cnn}^{1}、F_{cnn}^{2} 捕捉像素级局部细节，深层特征 F_{cnn}^{3}、F_{cnn}^{4} 则提取更抽象的语义表征。WHFC 模块通过二维离散 Haar 小波变换，提取图像的高频边缘信息，记为 F_{hfc}。与此同时，ViT 分支将输入图像转换为补丁嵌入 $F_{pe}^{i}\left(i=1,2,\cdots,N\right)$，保留图像的全局结构与上下文信息。本文提出的 MEAA 模块对上述三类任务专属特征(F_{cnn}^{i}、F_{hfc}、F_{pe}^{i})进行对齐与自适应融合，生成边缘感知提示嵌入 P_i，并将其逐阶段注入 ViT 分支，引导特征学习过程，强化模型对局部边缘细节与全局语义上下文的捕捉能力。经过完整编码器处理后，经自注意力(Self-attention, SA)精炼的 CNN 特征 $F_c^{'}$ 与 ViT 特征 $F_v^{'}$ 先通过逐元素相加完成初步融合，再经自注意力层进一步优化。随后，该融合特征与由点击提示经卷积层生成的密集图视觉提示进行逐元素相加，得到最终提示增强特征图，最后将其输入解码器通过上采样与特征恢复生成息肉分割掩码。

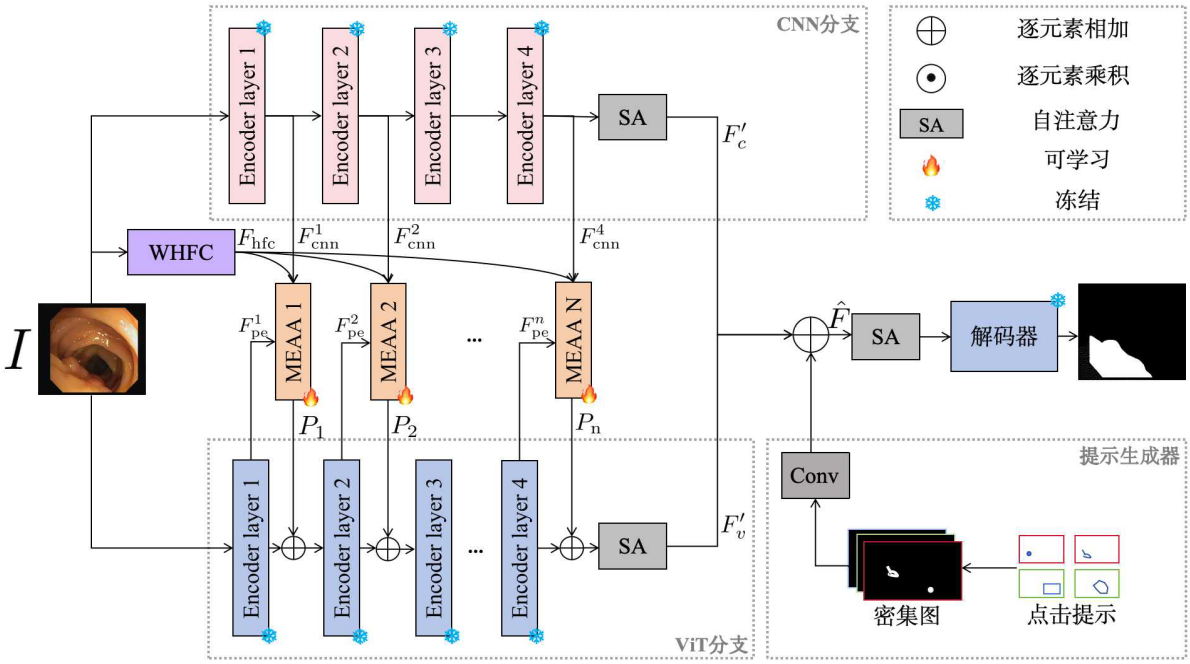

Figure 1. Overall model structure
图 1. 模型整体结构

2.1. MEAA 模块

2.1.1. 三种输入流

MEAA 模块通过整合三类异构特征流，实现对息肉区域更全面的特征表示与理解。在 CNN 特征流中，针对息肉纹理和语义特征提取需求，将 CNN 提取的多级特征分为浅层特征 $F_{cnn}^{1,2}$ 与深层特征 $F_{cnn}^{3,4}$ 并进行分层处理。其中，浅层特征经由浅层适配器细化，适配器由 1×1 卷积与非线性激活函数组成，用于精准捕捉黏膜纹理等细粒度局部细节；深层特征则通过高层适配器加工，该适配器由 3×3 卷积与残差连接组成，可输出息肉与周围组织的抽象语义表示，兼顾局部细节与全局结构的建模需求。高频边缘特征流旨在显式聚合息肉边界线索，通过二维离散 Haar 小波变换(Discrete Wavelet Transform, DWT)提取高频边缘特征。为模型提供息肉分割任务专属的边缘先验信息，具体实现过程见 1.2 节。ViT 特征流则依托补丁

嵌入 $F_{cnn}^{1,2}$ 捕捉全局上下文信息，通过自注意力机制有效刻画息肉病灶与周围正常组织间的长程空间依赖关系，弥补 CNN 在全局语义建模上的不足，为多模态特征融合提供全局语义支撑。

进一步地，浅层/深层特征的分离并非简单的层级拆分，而是针对息肉分割中“边缘细节定位”与“病灶语义判别”两类需求所作的功能解耦：浅层特征保留较高空间分辨率与纹理响应，更有利于刻画息肉边界、黏膜纹理及微小凸起等细粒度结构；深层特征具有更大的感受野和更强的语义抽象能力，更适合表征息肉与周围正常组织之间的语义差异。基于此，本文分别采用轻量浅层适配器与高层语义适配器对两类特征进行差异化处理，再在 MEAA 中与 WHFC 边缘先验及 ViT 全局语义共同融合，以避免直接混合多级 CNN 特征所导致的细节淹没或语义冲突。

2.1.2. MEAA 结构设计

在特征提取阶段之后，三类的特征图 F_{pe}、F_{cnn} 和 F_{hfc} 被送入 MEAA 进行多模态特征融合，其流程如图 2 所示。首先，对各特征分支分别执行全局平均池化(Global Average Pooling)以获得对应特征的通道平均权重，随后通过 Sigmoid 激活函数将其归一化至[0,1]区间，实现通道维度的特征权重分配。与此同时，各尺度特征经 1×1 卷积层完成通道压缩，生成单通道特征图；对压缩后的单通道特征图分别执行全局最大池化(Global Max Pooling)与全局平均池化，将两路输出沿通道维度拼接后，通过 7×7 卷积生成空间权重图，实现空间维度的特征权重分配。该空间权重同样经 Sigmoid 激活函数归一化至[0,1]区间，并与相应的通道权重逐元素相乘，最终得到通道–空间双重注意力权重。需要说明的是，图 2 中双注意力并非直接沿用 CBAM 的串行“通道注意力–空间注意力”重标定方式，而是针对多模态输入场景进行了改进：一方面，本文分别为 CNN、WHFC 与 ViT 三个分支独立构建通道与空间响应，使不同模态先在各自特征域内完成显著性估计；另一方面，在得到各分支的双注意力权重后，再通过 softmax 在模态维度上进行竞争式归一化，从而显式建模三类特征对当前分割任务的相对贡献。相比 CBAM 面向单一特征图的顺序增强方式，该设计更适合处理边缘先验、局部纹理与全局语义之间统计分布差异较大的异构特征，可在抑制冗余响应的同时保留互补信息。具体计算过程如下：

$$O_{cnn}=C_{7\times7}(Cat\left[GMP\left(C_{1\times1}\left(F_{cnn}^{i}\right),GAP\left(C_{1\times1}\left(F_{cnn}^{i}\right)\right)\right]\right)\odot\sigma\left(GAP\left(F_{cnn}^{i}\right)\right) \tag{1}$$

$$O_{hfc}=C_{7\times7}(Cat\left[GMP\left(C_{1\times1}\left(F_{hfc}\right),GAP\left(C_{1\times1}\left(F_{hfc}\right)\right)\right]\right)\odot\sigma\left(GAP\left(F_{hfc}\right)\right) \tag{2}$$

$$O_{pe}=C_{7\times7}(Cat\left[GMP\left(C_{1\times1}\left(F_{pe}^{i}\right),GAP\left(C_{1\times1}\left(F_{pe}^{i}\right)\right)\right]\right)\odot\sigma\left(GAP\left(F_{pe}^{i}\right)\right) \tag{3}$$

随后，在三个特征分支上应用 softmax 函数对双重注意力权重进行归一化。这些归一化权重通过逐元素乘法对各输入特征进行重加权，再将重加权后的特征相加产生多模态融合表示。最后，融合特征依次通过两个多层感知机(*MLP*)以及 *MLP* 内部的 GELU 激活函数，生成提示 $P_i\left(i=1,2,\cdots,n\right)$：

此外，图 2 中标记为 MLP_{tune} 与 MLP_{up} 的两个 *MLP* 具有不同的参数共享策略：其中，MLP_{tune} 采用 layer-unshared 方式，即在不同阶段的 MEAA 模块中分别独立设置参数，以适应不同编码层对局部纹理、边缘信息及语义上下文的差异化建模需求；MLP_{up} 采用 layer-shared 方式，即在各阶段的 MEAA 中共享同一组参数，其目的在于以统一映射规则完成融合特征到 ViT 提示嵌入空间的维度提升与分布对齐。换言之，文中“shared/unshared”描述的是模型整体不同阶段之间的参数共享关系，而非单个 MEAA 模块内部不同计算步骤之间的共享关系。

$$F_i=Softmax\left(\left(O_{cnn}\odot F_{cnn}^{i}\right)\oplus\left(O_{hfc}\odot F_{hfc}\right)\oplus\left(O_{pe}\odot F_{pe}^{i}\right)\right) \tag{4}$$

$$P_i=MLP_{up}\left(GELU\left(MLP_{tune}^{i}\left(F_i\right)\right)\right) \tag{5}$$

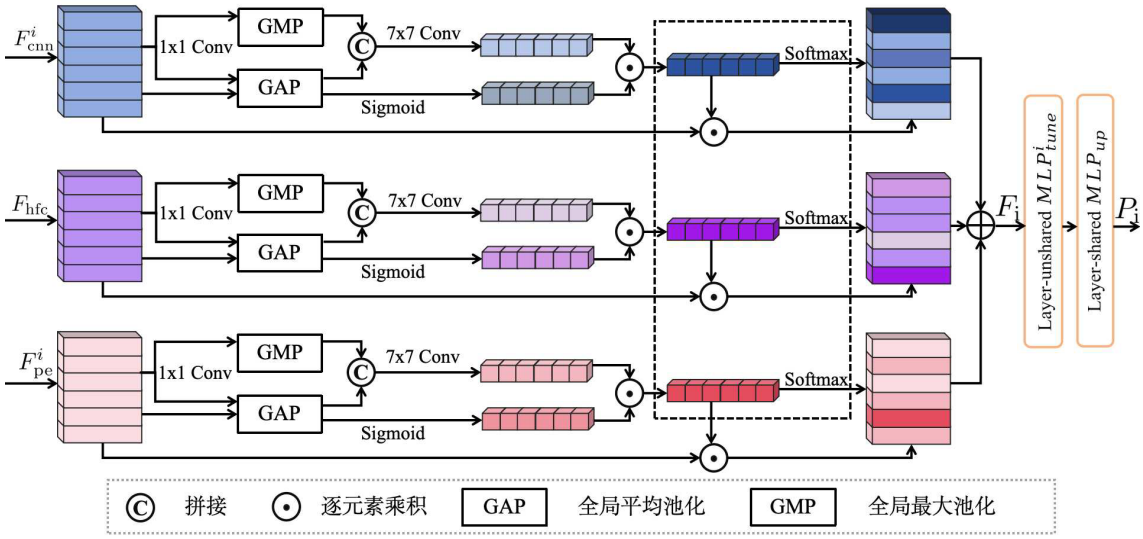

Figure 2. Structure of MEAA module
图 2. MEAA 模型结构

其中 MLP^i_{tune} 为特征调优多层感知机，实现融合特征的精细化表征； MLP_{up} 为维度提升多层感知机，实现融合特征与 ViT 分支嵌入维度的对齐。具体而言，MLP_{tune} 主要承担融合特征细化与任务适配的功能，因此保持阶段间参数独立，有助于不同层级学习与其感受野相匹配的调制方式； MLP_{up} 则侧重于将融合表示稳定映射到与 ViT 分支一致的嵌入维度，采用跨阶段共享参数能够减少额外参数量，并增强不同阶段提示表征的一致性。

2.2. 基于小波变换的高频分量模块设计

在息肉分割任务中，精准捕捉息肉的边缘结构是实现病灶边界清晰勾勒的关键。为解决现有方法边缘细节捕捉不足的问题，本文引入基于二维离散小波变换(Discrete Wavelet Transform, DWT)的高频边缘特征提取方法(WHFC)，实现对息肉边缘细节的有效捕捉，整体流程如图 3 所示。该模块选用 Haar 小波构建滤波器组，通过低通和高通滤波器协同作用，精准提取息肉边缘高频特征。选择 Haar 小波的原因主要在于：其基函数结构简单、局部性强、计算开销低，能够以较小的额外代价突出图像灰度突变位置，对息肉边界这类高频不连续区域具有更直接的响应；同时，Haar 小波不依赖较长支撑域，能在保留边缘尖锐性的同时减少过度平滑现象，更符合息肉分割中对清晰边界先验的需求。

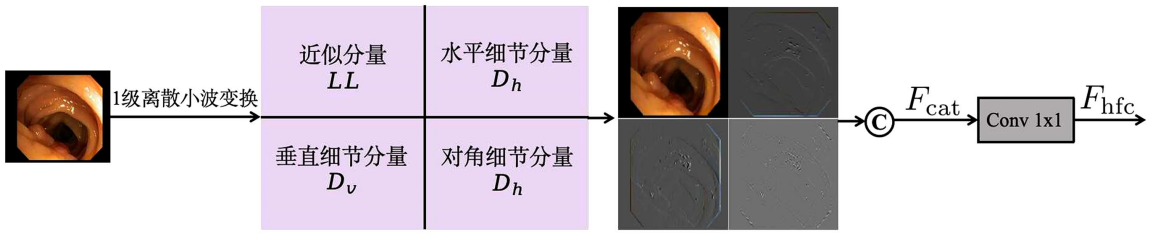

Figure 3. The process to generate high-frequency edge features
图 3. 生成高频边缘特征的过程

输入图像 $I \in \mathbb{R}^{H\times W\times 3}$ 的每个颜色通道独立执行二维离散小波变换，分解得到四个分量，分别为低频近似分量 LL 与水平、垂直、对角三个高频分量 D_h、D_v、D_d。本研究选用 Haar 小波构建滤波器组，其低

通滤波器(近似滤波器) $h_l = \frac{1}{\sqrt{2}}[1,1]$，高通滤波器(细节滤波器) $h_h = \frac{1}{\sqrt{2}}[1,-1]$，通过高低通滤波器的协同分解实现图像高频边缘特征与低频整体特征的有效分离。小波分解过程由卷积运算与下采样共同完成，各分量的具体生成方式如下：

$$\begin{cases} \text{低频近似分量：} LL = \left(I \times h_l \downarrow_2\right) \times h_l \downarrow_2 \\ \text{水平近似分量：} D_h = \left(I \times h_l \downarrow_2\right) \times h_h \downarrow_2 \\ \text{垂直近似分量：} D_v = \left(I \times h_h \downarrow_2\right) \times h_l \downarrow_2 \\ \text{对角近似分量：} D_d = \left(I \times h_h \downarrow_2\right) \times h_h \downarrow_2 \end{cases} \tag{6}$$

其中$\downarrow_2$表示因子为 2 的下采样操作，仅保留图像的关键特征信息。

随后仅保留水平、垂直、对角三个高频分量 D_h、D_v、$D_d \in \mathbb{R}^{\frac{H}{2}\times\frac{W}{2}\times 1}$，将这三个方向的高频分量沿通道维度进行拼接，构建得到高维特征图 F_{cat}，其中，每个高频分量 $D_{h,v,d}^{RGB}$ 均包含来自 *RGB* 三通道的高频信息(即分别堆叠水平、垂直和对角方向 *R*、*G*、*B* 通道高频特征 D_h^R、D_h^G 与 D_h^B)，完整保留息肉边缘的色彩与纹理信息。

随后通过可学习的 1 × 1 卷积层对高维边缘特征图进行通道降维处理，同时完成更高层级的边缘特征表征提取，最终得到与 ViT 中间层嵌入维度完全对齐的高频边界特征表示 F_{hfc}，该特征可有效捕捉图像中的息肉边缘高频细节，为后续多模态特征融合及息肉分割任务提供丰富的边界特征信息。其数学表达为：

$$F_{hfc} = Conv_{1\times1}\left(F_{cat}\right) \tag{7}$$

3. 实验结果与分析

3.1. 数据集

本文实验参照 PraNet [9]的实验方法展开，选用 Kvasir-SEG [19]、CVC-ClinicDB [20]、CVC-ColonDB [21]、ETIS [22]、EndoScene [23]五款公开的息肉分割数据集完成模型性能验证，数据集的训练集与测试集划分方式保持与 PraNet 一致：训练集由 Kvasir-SEG 的 900 幅图像与 CVC-ClinicDB 的 550 幅图像组成；测试集包含 Kvasir-SEG 100 幅、CVC-ClinicDB 62 幅、CVC-ColonDB 380 幅、EndoScene 60 幅及 ETIS 196 幅图像。

3.2. 实验设置

模型的训练与推理过程均沿用 SegNext 的提示模块设置，将点击位置编码至固定半径的圆盘映射图中，圆盘半径设为 5。实验基于 40 GB 显存的 A100 GPU 完成，模型优化器选用 Adam，初始学习率设置为 0.01，采用阶梯式学习率衰减策略，每训练 50 个轮次将学习率衰减至原有值的 1/10，总训练轮次为 200，批次尺寸设为 2。模型性能评估采用医学图像分割领域的两项常用评价指标，即 Dice 系数与交并比(IoU)。针对数据集内图像分辨率不一致的问题，训练阶段先将图像最长边缩放至 1024 像素，再通过零填充操作将图像统一调整为 1024 × 1024 的尺寸；推理阶段则直接将测试图像缩放至 1024 × 1024 的尺寸，不进行填充处理，避免填充区域对分割结果造成干扰。

本文以点击交互作为定量对比的默认交互方式，点击操作依据前一次预测结果的误差实现自动模拟。首次点击时，模型无前期分割结果作为参考，即初始分割结果为空，模拟临床实际的分割场景。

3.3. 基线模型

为全面验证所提方法的分割性能，本文选取多项当前主流息肉分割方法及基于 SAM 的相关方法进行公平对比，涵盖基于 CNN、ViT 及 SAM 提示工程的各类先进模型。对比方法具体包括：传统 CNN 类方法 UNet [1]、PraNet [9]；CNN 与 Transformer 融合类方法 SANet [24]、MSNet [25]、LDNet [26]、SSFormer [27]、Polyp-PVT [28]；基于 SAM 的改进方法 SAM-H [29]、SAM-L [29]、SAMAdapter [30]、ASPS [31]、MedSAM [32]，同时将原始 SegNext 在 1 次、2 次交互提示下的性能作为基线参考，所有对比模型均在相同的实验数据集与训练设置下完成训练与测试。

3.4. 定量比较与分析

本文将所提方法与当前主流息肉分割方法、各类基于 SAM 的分割方法的性能进行对比，详细结果如表 1 所示。表中加粗表示该指标下的最优性能，下划线表示次优性能。

实验结果表明，所提方法在 CVC-ClinicDB 数据集上保持了与最优基线模型(ASPS)相当的性能，而在 CVC-ColonDB、ETIS、Kvasir 及 EndoScene 数据集上均取得了明显的性能改善，展现出良好的泛化能力，适用于不同类型的息肉分割场景。ASPS 模型在 CVC-ClinicDB 与 EndoScene 数据集上表现出较优的性能，其在 CVC-ClinicDB 上的 Dice 与 IoU 分别达到 0.952 与 0.907，在 EndoScene 上的 Dice 与 IoU 分别为 0.927 与 0.883，综合平均 Dice 系数与 IoU 分别为 0.889 和 0.819。本文所提方法在不同交互提示设置下，整体分割性能显著优于对比方法：在无交互提示的(0-m)设置下，模型平均 Dice 系数与 IoU 为 0.854 和 0.781；在(1-m)单点击提示设置下提升至 0.910 和 0.858；在(2-m)双点击提示设置下进一步达到 0.935 和 0.895。具体来看，在 CVC-ClinicDB 数据集上，所提方法(2-m)设置下的 Dice (0.948)与 IoU (0.910)略低于 ASPS 模型(Dice 0.952, IoU 0.907)，但在其余 4 个数据集上均实现了超越；在 CVC-ColonDB、ETIS 等难度更高的数据集上，性能提升尤为显著，充分验证了本模型的有效性与优越性。

相较于各类基于 SAM 的分割方法，本文所提模型在未引入点击提示(0-m)的情况下，性能已优于多数对比模型，且在 CVC-ColonDB、ETIS 与 Kvasir 数据集上的表现超越 ASPS 模型。在(1-m)单点击提示的实验设置下，本文模型在除 CVC-ClinicDB 外的其余所有数据集上，分割性能均显著优于对比方法。在双点击提示(2-m)下，模型可进一步利用交互信息细化息肉边界与模糊区域，各项指标提升幅度更为明显，充分体现所提方法在不同交互强度下均具备稳定且优异的泛化能力。同时实验发现，SAMUS 与 ASPS 模型因引入了 CNN 辅助分支，也取得了较好的分割效果，这一结果也进一步验证了本文所设计 CNN 分支在息肉分割任务中的有效性与合理性。

Table 1. Quantitative comparison of methods across five publicly available polyp datasets
表 1. 五个公开息肉数据集上的最先进方法定量比较

模型	CVC-ClinicDB		Kvasir		CVC-ColonDB		ETIS		EndoScene	
	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
UNet	0.823	0.755	0.818	0.746	0.504	0.436	0.398	0.335	0.710	0.627
PraNet	0.899	0.849	0.898	0.840	0.709	0.640	0.628	0.567	0.871	0.797
MSNet	0.916	0.859	0.904	0.847	0.752	0.669	0.750	0.654	0.888	0.815
LDNet	0.932	0.872	0.912	0.855	0.794	0.715	0.778	0.707	0.893	0.826
SSFormer	0.906	0.855	0.917	0.864	0.802	0.721	0.796	0.720	0.895	0.827
Polyp-PVT	0.937	0.889	0.917	0.864	0.808	0.727	0.787	0.706	0.900	0.833
SAM-H	0.547	0.500	0.778	0.707	0.441	0.396	0.517	0.447	0.651	0.606

续表

SAM-L	0.579	0.526	0.782	0.710	0.468	0.422	0.551	0.507	0.726	0.676
SAM-Adapter	0.774	0.673	0.847	0.763	0.671	0.568	0.590	0.476	0.815	0.725
SAMed	0.404	0.273	0.459	0.300	0.199	0.115	0.212	0.126	0.332	0.202
SAMUS	0.900	0.821	0.859	0.763	0.731	0.597	0.750	0.618	0.859	0.760
ASPS	**0.952**	**0.907**	0.914	0.848	0.792	0.694	0.856	0.764	**0.927**	**0.883**
MedSAM	0.867	0.803	0.862	0.795	0.734	0.651	0.687	0.604	0.870	0.798
SegNext (1-m)	0.436	0.357	0.721	0.640	0.417	0.335	0.368	0.308	0.603	0.525
SegNext (2-m)	0.678	0.532	0.850	0.783	0.680	0.579	0.615	0.530	0.803	0.725
Ours (0-m)	0.865	0.802	0.901	0.842	0.821	0.736	0.819	0.735	0.863	0.792
Ours (1-m)	0.935	0.884	0.936	0.898	0.883	0.792	0.897	0.841	0.902	0.864
Ours (2-m)	0.948	0.910	**0.955**	**0.926**	**0.924**	**0.873**	**0.949**	**0.887**	0.917	0.879

3.5. 定性比较与分析

图 4 展示了本文所提模型与 PraNet、SAMAdapter、MedSAM、SAMUS、ASPS 等主流方法的息肉分割结果可视化对比。针对第一幅含微小息肉目标的测试图像，PraNet 与 SAMAdapter 方法未能实现对息肉有效区域的完整分割，凸显出两类方法在小目标息肉检测与分割任务中的局限性。而 MedSAM、SAMUS 与 ASPS 模型的分割结果存在边缘定位不准的问题，与之相比，本文所提模型的息肉边缘分割精度表现优异。该方法通过密集图对视觉提示进行表征，能够在保留息肉整体结构的前提下，精准定位并勾勒出息肉的细微边缘轮廓。

针对第二幅含低对比度不规则形状息肉的测试图像，PraNet、SAMAdapter、MedSAM 与 SAMUS 模型均未能实现息肉有效区域的完整分割，ASPS 模型虽能分割整体区域，但仍存在息肉边缘定位不准的问题。而本文所提模型在 1-m 单提示设置下，即可精准分割出息肉的有效区域；且随着提示点数量的增加，模型对息肉边界的感知能力进一步提升，有效解决了息肉边缘分割模糊的问题。

第三幅测试图像中的息肉区域存在对比度复杂、光照变化显著的问题。其他对比模型仅能成功分割出息肉中与背景对比度较高的上部区域，对于息肉下部区域，均出现了不同程度的边界模糊问题，且未能完成有效区域的完整分割。与之相比，本文所提模型即便在这类复杂背景下，仍能精准识别并完整分割出整个息肉区域，证明了模型在复杂场景下的鲁棒性。整体而言，所提方法在息肉边缘细节、小目标息肉、低对比度息肉分割上均表现出显著优势，能够更好地满足临床对息肉精准分割的需求。

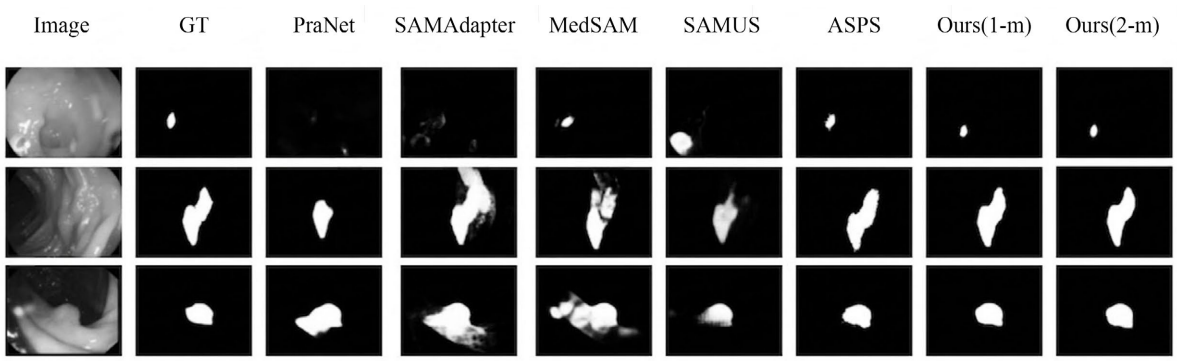

Figure 4. Qualitative comparison of segmentation results
图 4. 分割结果的可视化对比

3.6. 消融实验

3.6.1. 模块消融分析

为验证所引入的 CNN 分支与 MEAA 模块的有效性，本文设计了相应的消融实验。消融实验未引入交互提示，与主实验 0-m 无提示设置保持一致，以 SegNext 中的 ViT-B 作为基准模型主干网络，实验结果如表 2 所示。当模型仅保留 ViT-B 主干网络时，分割性能较差，平均 Dice 系数与 IoU 仅为 0.565 和 0.493，证明单一 ViT 主干网络难以有效完成息肉分割任务；加入 CNN 分支后，模型的分割性能得到大幅提升，平均 Dice 系数与 IoU 提升至 0.829 和 0.762，证明 CNN 分支能够有效提取息肉的局部细节特征，弥补 ViT 在局部特征建模上的不足；在加入 CNN 分支的基础上进一步加入 MEAA 模块后，模型性能得到进一步提升，平均 Dice 系数与 IoU 达到 0.855 和 0.782，证明 MEAA 模块能够有效实现多模态特征的自适应融合，充分挖掘异构特征间的互补价值，提升模型的分割性能。

Table 2. Ablation experiments for core components
表 2. 核心模块消融实验结果

CNN	MEAA	CVC-ClinicDB		Kvasir		CVC-ColonDB		ETIS		EndoScene	
		Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
		0.574	0.521	0.603	0.542	0.518	0.451	0.542	0.459	0.586	0.492
√		0.841	0.777	0.866	0.809	0.789	0.724	0.798	0.728	0.850	0.771
√	√	0.872	0.805	0.901	0.842	0.821	0.736	0.819	0.735	0.863	0.792

3.6.2. MEAA 模块消融实验分析

为验证 MEAA 模块中三类输入特征流的有效性，本文对其输入的 3 个关键组件进行了消融实验，分别为 CNN 分支特征 F_{cnn}、WHFC 模块特征 F_{cnn} 与 ViT 补丁嵌入特征 F_{pe}。实验以仅包含 CNN 分支的模型作为基准模型，实验结果如表 3 所示。

从实验结果可以看出，单独加入任意一类特征流均能提升模型的分割性能，证明三类特征流在息肉分割任务中均具有重要作用，能够提供互补的特征信息；同时加入两类特征流时，模型性能得到进一步提升，而同时加入三类特征流时，模型取得了最优的分割性能，平均 Dice 系数与 IoU 达到 0.870 和 0.801。这一结果进一步表明 MEAA 模块能够有效利用不同特征之间的互补性，通过 - 空间双注意力机制实现多模态特征的自适应融合，显著提升模型的分割效果。

Table 3. Ablation experiments for feature streams in MEAA
表 3. MEAA 模块内部特征流消融实验结果

MEAA			CVC-ClinicDB		Kvasir		CVC-ColonDB		ETIS		EndoScene	
F_{cnn}	F_{hfc}	F_{pe}	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
			0.863	0.802	0.894	0.832	0.816	0.748	0.827	0.744	0.871	0.791
√			0.876	0.813	0.904	0.852	0.830	0.749	0.827	0.748	0.868	0.801
	√		0.872	0.810	0.901	0.849	0.825	0.746	0.824	0.745	0.864	0.797
		√	0.874	0.812	0.902	0.850	0.828	0.747	0.830	0.746	0.866	0.799
√	√		0.873	0.819	0.912	0.855	0.833	0.742	0.830	0.750	0.870	0.804
√		√	0.875	0.821	0.910	0.854	0.835	0.752	0.831	0.751	0.872	0.805
	√	√	0.878	0.817	0.917	0.853	0.832	0.750	0.829	0.750	0.871	0.803
√	√	√	0.885	0.820	0.913	0.856	0.836	0.762	0.833	0.755	0.881	0.811

3.6.3. WHFC 模块消融实验分析

为验证 WHFC 模块在提升边缘感知与边界分割精度方面的有效性，本文对不同边缘提取策略进行了对比实验。以原始 WHFC 结构为基准，分别将 Haar 小波替换为 Daubechies、Symlet 及 Coiflet 小波，并对比了 Haar 小波在不同分解层数下的性能影响。所有实验均在相同参数设置下训练，并在 CVC-ColonDB 与 ETIS 数据集上进行评估。除 WHFC (5 层)采用 5 级小波分解外，其余小波基均统一使用 3 级分解。为进一步回应 WHFC 中小波基选择的合理性，本文补充对比了不同小波基及不同分解层数下的分割性能，并将该实验作为 Haar 小波选择依据的实证支撑。

由表 4 可知，将 Haar 小波替换为 Daubechies、Symlet 或 Coiflet 小波后，模型性能均出现不同程度下降。这表明 Haar 小波相较于其他小波基具有更优异的边缘敏感特性，更适用于医学图像中的细边缘检测任务。此外，实验结果表明，采用 5 级离散小波变换分解会引入额外的低频干扰，反而导致模型性能降低，而 3 级分解能够在边缘特征提取与计算复杂度之间取得最优平衡。上述结果进一步验证了 WHFC 模块设计的有效性与优越性。从机理上看，Daubechies、Symlet 与 Coiflet 等小波虽然具有更强的平滑性与更长的支撑长度，但在息肉边缘较细、局部灰度突变明显的场景中，容易在分解过程中引入更宽的响应范围，削弱边界定位的敏锐性；相较之下，Haar 小波的分段常数基更利于保留突变边缘，因此在本任务中取得了更优结果。结合表 4 的实验结果可知，本文对 Haar 小波与 3-level 分解深度的选择具有一定的理论依据与实验依据。

Table 4. Ablation experiments for WHFC module
表 4. WHFC 模块消融实验结果

模型	CVC-ColonDB		ETIS	
	Dice	IoU	Dice	IoU
Daubechies	0.740	0.704	0.717	0.644
Symlet	0.804	0.729	0.744	0.694
Coiflet	0.745	0.694	0.742	0.689
WHFC (5-level)	0.819	0.729	0.791	0.697
WHFC (3-level)	0.849	0.749	0.842	0.769

4. 结论

本文针对结肠息肉精准分割中特征融合不充分、边缘细节丢失、泛化能力不足的问题，同时弥补 SegNext 框架的针对性缺陷，提出基于该框架的多模态边缘感知适配器息肉分割方法。该方法构建了 CNN、WHFC 与 ViT 多分支特征提取结构，分别提取息肉局部纹理、边缘先验及全局语义特征；通过 MEAA 模块实现异构特征自适应加权融合，挖掘局部与全局特征的互补价值；结合 SegNext 密集图提示策略，双重强化空间细节与高层语义，提升模型在复杂医学图像中的分割精度与鲁棒性。

实验结果显示，所提方法在 5 个公开数据集上性能优于主流方法，无提示、两次连续提示场景下平均 Dice/IoU 分别达 0.854/0.781、0.935/0.895，具备临床实时应用潜力，为结直肠癌早期诊断提供高效辅助支撑。

本方法仍有局限：训练需依托高性能 GPU，暂未适配边缘设备；仅在息肉分割任务中验证，未拓展至其他医学图像分割场景。未来将优化模型架构，通过量化、剪枝降低计算复杂度，适配资源受限环境；同时拓展至更多医学分割任务，融合临床先验与医生反馈，提升模型实用性与可解释性，为智能医疗辅助诊断技术发展提供新方向。

基金项目

国家重点研发计划项目(2021YFF0600605)。

参考文献

[1] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[2] Oktay, O., Schlemper, J., Le Folgoc, L., *et al*. (2018) Attention U-Net: Learning Where to Look for the Pancreas. arXiv:1804.03999. https://arxiv.org/abs/1804.03999

[3] Li, L., Verma, M., Nakashima, Y., Nagahara, H. and Kawasaki, R. (2020) IterNet: Retinal Image Segmentation Utilizing Structural Redundancy in Vessel Networks. 2020 *IEEE Winter Conference on Applications of Computer Vision* (*WACV*), Snowmass, 1-5 March 2020, 3656-3665. https://doi.org/10.1109/wacv45572.2020.9093621

[4] Alom, M.Z., Yakopcic, C., Taha, T.M. and Asari, V.K. (2018) Nuclei Segmentation with Recurrent Residual Convolutional Neural Networks Based U-Net (R2U-Net). *NAECON* 2018-*IEEE National Aerospace and Electronics Conference*, Dayton, 23-26 July 2018, 228-233. https://doi.org/10.1109/naecon.2018.8556686

[5] Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., *et al*. (2019) Ce-Net: Context Encoder Network for 2D Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, **38**, 2281-2292. https://doi.org/10.1109/tmi.2019.2903562

[6] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N. and Liang, J. (2020) UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging*, **39**, 1856-1867. https://doi.org/10.1109/tmi.2019.2959609

[7] Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., Lange, T.D., Halvorsen, P., *et al*. (2019) ResUNet++: An Advanced Architecture for Medical Image Segmentation. 2019 *IEEE International Symposium on Multimedia* (*ISM*), San Diego, 9-11 December 2019, 225-255. https://doi.org/10.1109/ism46123.2019.00049

[8] Tomar, N.K., Jha, D., Ali, S., Johansen, H.D., Johansen, D., Riegler, M.A., *et al*. (2021) DDANet: Dual Decoder Attention Network for Automatic Polyp Segmentation. In: Del Bimbo, A., *et al*., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 307-314. https://doi.org/10.1007/978-3-030-68793-9_23

[9] Fan, D., Ji, G., Zhou, T., Chen, G., Fu, H., Shen, J., *et al*. (2020) PraNet: Parallel Reverse Attention Network for Polyp Segmentation. In: Martel, A.L., *et al*., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 263-273. https://doi.org/10.1007/978-3-030-59725-2_26

[10] Zhang, Y., Liu, H. and Hu, Q. (2021) TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation. In: de Bruijne, M., *et al*., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 14-24. https://doi.org/10.1007/978-3-030-87193-2_2

[11] Fan, X., Zhou, J., Jiang, X., Xin, M. and Hou, L. (2024) CSAP-UNet: Convolution and Self-Attention Paralleling Network for Medical Image Segmentation with Edge Enhancement. *Computers in Biology and Medicine*, **172**, Article 108265. https://doi.org/10.1016/j.compbiomed.2024.108265

[12] Yuan, F., Zhang, Z. and Fang, Z. (2023) An Effective CNN and Transformer Complementary Network for Medical Image Segmentation. *Pattern Recognition*, **136**, Article 109228. https://doi.org/10.1016/j.patcog.2022.109228

[13] Chen, J., Lu, Y., Yu, Q., *et al*. (2021) TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv:2102.04306. https://arxiv.org/abs/2102.04306

[14] Zhang, K. and Liu, D. (2023) Customized Segment Anything Model for Medical Image Segmentation. arXiv:2304.13785. https://arxiv.org/abs/2304.13785

[15] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., *et al*. (2023) Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. In: Karlinsky, L., Michaeli, T. and Nishino, K., Eds., *Lecture Notes in Computer Science*, Springer, 205-218. https://doi.org/10.1007/978-3-031-25066-8_9

[16] Roy, S., Wald, T., Koehler, G., *et al*. (2023) SAM.MD: Zero-Shot Medical Image Segmentation Capabilities of the Segment Anything Model. arXiv:2304.05396. https://arxiv.org/abs/2304.05396

[17] Bui, N., Hoang, D., Tran, M., Doretto, G., Adjeroh, D., Patel, B., *et al*. (2024) SAM3D: Segment Anything Model in Volumetric Medical Images. 2024 *IEEE International Symposium on Biomedical Imaging* (*ISBI*), Athens, 27-30 May 2024, 1-4. https://doi.org/10.1109/isbi56570.2024.10635844

[18] Liu, Q., Cho, J., Bansal, M. and Niethammer, M. (2024) Rethinking Interactive Image Segmentation with Low Latency, High Quality, and Diverse Prompts. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 16-22 June 2024, 3773-3782. https://doi.org/10.1109/cvpr52733.2024.00362

[19] Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., de Lange, T., Johansen, D., *et al*. (2019) Kvasir-SEG: A Segmented Polyp Dataset. In: Ro, Y., *et al*., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 451-462. https://doi.org/10.1007/978-3-030-37734-2_37

[20] Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C. and Vilariño, F. (2015) WM-DOVA Maps for Accurate Polyp Highlighting in Colonoscopy: Validation vs. Saliency Maps from Physicians. *Computerized Medical Imaging and Graphics*, **43**, 99-111. https://doi.org/10.1016/j.compmedimag.2015.02.007

[21] Tajbakhsh, N., Gurudu, S.R. and Liang, J. (2016) Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information. *IEEE Transactions on Medical Imaging*, **35**, 630-644. https://doi.org/10.1109/tmi.2015.2487997

[22] Silva, J., Histace, A., Romain, O., Dray, X. and Granado, B. (2014) Toward Embedded Detection of Polyps in WCE Images for Early Diagnosis of Colorectal Cancer. *International Journal of Computer Assisted Radiology and Surgery*, **9**, 283-293. https://doi.org/10.1007/s11548-013-0926-3

[23] Vazquez, D., Bernal, J., Sanchez, F.J., *et al*. (2016) A Benchmark for Endoluminal Scene Segmentation of Colonoscopy Images. arXiv:1612.00799. https://arxiv.org/abs/1612.00799

[24] Fan, H. and Ling, H. (2017) SANet: Structure-Aware Network for Visual Tracking. 2017 *IEEE Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), Honolulu, 21-26 July 2017, 1-8. https://doi.org/10.1109/cvprw.2017.275

[25] Zhu, X., Liang, J. and Hauptmann, A.G. (2020) MSNet: A Multilevel Instance Segmentation Network for Natural Disaster Damage Assessment in Aerial Videos. arXiv:2006.16479. https://arxiv.org/abs/2006.16479

[26] Zhang, R., Lai, P., Wan, X., *et al*. (2023) Lesion-Aware Dynamic Kernel for Polyp Segmentation. arXiv:2301.04904. https://arxiv.org/abs/2301.04904

[27] Shi, W., Xu, J. and Gao, P. (2022) SSformer: A Lightweight Transformer for Semantic Segmentation. 2022 *IEEE* 24*th International Workshop on Multimedia Signal Processing* (*MMSP*), Shanghai, 26-28 September 2022, 1-5. https://doi.org/10.1109/mmsp55362.2022.9949177

[28] Dong, B., Wang, W., Fan, D., Li, J., Fu, H. and Shao, L. (2023) Polyp-PVT: Polyp Segmentation with Pyramid Vision Transformers. *CAAI Artificial Intelligence Research*, **2**, 9150015. https://doi.org/10.26599/air.2023.9150015

[29] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., *et al*. (2023) Segment Anything. 2023 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Paris, 1-6 October 2023, 3992-4003. https://doi.org/10.1109/iccv51070.2023.00371

[30] Chen, T., Zhu, L., Ding, C., Cao, R., Wang, Y., Zhang, S., *et al*. (2023) SAM-Adapter: Adapting Segment Anything in Underperformed Scenes. 2023 *IEEE*/*CVF International Conference on Computer Vision Workshops* (*ICCVW*), Paris, 2-6 October 2023, 3359-3367. https://doi.org/10.1109/iccvw60793.2023.00361

[31] Li, H., Zhang, D., Yao, J., Han, L., Li, Z. and Han, J. (2024) ASPS: Augmented Segment Anything Model for Polyp Segmentation. In: Linguraru, M.G., *et al*., Eds., *Lecture Notes in Computer Science*, Springer, 118-128. https://doi.org/10.1007/978-3-031-72114-4_12

[32] Ma, J., He, Y., Li, F., Han, L., You, C. and Wang, B. (2024) Segment Anything in Medical Images. *Nature Communications*, **15**, Article No. 654. https://doi.org/10.1038/s41467-024-44824-z